DOCUMENT RESUME

ED 061 805 FL 003 009

AUTHOR Upshur, John A.
TITLE A Search for New Reading Tests.
PUB DATE 1 Mar 72
NOTE 12p.; Speech presented at the Sixth Annual TESOL Convention, Washington, D.C., March 1, 1972

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Cloze Procedure; *Comprehension; English (Second Language); Eye Movements; Grammar; *Information Processing; Language Instruction; Language Skills; Reading Ability; Reading Instruction; *Reading Tests; *Second Language Learning; Semantics; *Test Construction; Test Reliability; Visual Acuity

ABSTRACT

In contemporary reading comprehension tests, abilities which may have little to do with reading ability can have a great effect on test scores. Reading tests are not constructed so that they maximize abilities unique to reading. On the other hand, tests on other language skills do not minimize what is unique to reading. Comprehension is the process of getting the message, and a reading comprehension test should give information about that stage of the information processing sequence. The cloze technique, photography of eye movement, and miscue analysis can all be used to measure a smaller range of information-processing tasks than do the contemporary reading comprehension tests. (VM)

ED 061805

A SEARCH FOR NEW READING TESTS*

John A. Upshur

University of Michigan

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

TL 003 004

*Prepared for oral presentation at the TESOL Convention, Washington, March, 1972. Not for publication.

Permission is granted for listing and reproduction through ALCA/ERIC.

A SEARCH FOR NEW READING TESTS

I've sometimes wondered whether people other than conference organizers care about some of the topics that conference speakers are asked to talk about. This seemed to be a researchable question, so with myself as principal investigator, I conducted a small study. Subjects were teachers of modern foreign languages. The procedure was a structured interview beginning with the following question:

"Would you as a teacher like to see new reading ability tests?"

The data from one subject, who mistook me for a salesman, were discarded. Of the remaining five subjects, one said, "Yes," but could offer no reasons for wanting new tests; he was evidently trying to encourage me in whatever schemes I might have had in mind.

Two teachers answered, "No." One of them expressed a belief that any testing serves only to subvert the true meaning and process of education; the other stated that tests might be useful for some purposes but constitute a wasteful use of classroom time.

One subject wanted new tests because her students wanted them; they felt that taking tests helped them prepare for proficiency examinations which they would have to sit for later at the end of their course.

My last subject also wanted new tests. She gave as a reason that they might help her to teach better.

I thought to myself that the results of my study were not altogether discouraging. After all, one sixth of all teachers surveyed wanted a further source of information about students in order to improve instruction. And so, continuing my investigation, I asked this subject what kind of

information she wanted new tests to provide, but unfortunately she had no answer for me. I tried probing.

"Would you like to know the things that are going on in his head?"

"Yes, sure."

"Well, what sorts of things should be going on?"

"I don't know."

I tried another prompt.

"Suppose you knew that for some student, information was being replaced in short-term-memory before going into long-term-memory. What things, different from what you're doing now, would you do for that student?"

"I don't know."

Once again I felt pretty discouraged, but then she added, "But you know, if you could tell me something like that, I'd sure as hell pretty soon figure out or find out something to do about it."

In order to get a perspective on what's new, it might be in order first to review what is current and what is past. These temporal terms are relative to place, however. What has already passed in the United States, for example, is current practice in many parts of the world.

The two traditional techniques for testing reading are translation into the native language, and close analysis of the reading text in terms of surface grammar: identifying sentence subjects, naming verb tenses, and so forth. Each of these techniques reveals an implicit assumption about reading. First, that the purpose for reading a text is to obtain its meaning, a meaning which is fully translatable into the student's own language. The second assumption is that students can obtain meanings of texts when they know translation equivalents of content words and know, or can figure out,

structural relationships among the words and phrases of the component sentences of the text.

It is fairly easy at this remove to see a major problem with this testing procedure. It did not favor fluent reading so much as it did filling a text with interlinear notes. It also ran afoul of an emerging orthodoxy which proclaimed translation a "bad habit" to be avoided at all costs in language instruction.

Current testing practice reflects a variety of influences. Among these are certainly the four-skill model of language use, a behavioral approach to language instruction, and American school tests of reading.

The three textbooks on foreign language testing on the market today (Harris, 1969; Lado, 1961; Valette, 1967) devote a total of forty-nine pages to reading. Of these forty-nine pages more than a dozen are introduction or examples without comment. There is considerable overlap among the texts for the remaining pages; little is left on procedures for testing reading. But there is some.

A number of prereading test types are illustrated: matching letters, word matching, and matching written words to pictures or to their spoken forms. Grammar and vocabulary tests are suggested as fundamental to reading; tests of reading speed are proposed but not well explained. Most emphasis is placed upon the format of a passage to read followed by a set of four to six questions, usually of the multiple-choice type. This format is the one most used in standardized tests of first or second language reading.

In the passage-plus-questions format, comprehension is functionally defined as the ability to answer questions after exposure to a passage. Comprehension thus has the same status as does habit in stimulus-response

psychology. It is that unanalyzable construct intervening between text (stimulus) and question (response) which is inferred when correct answers occur with greater than chance probability.

One can vary stimulus, or passage, characteristics such as type-token ratio, syntactic complexity, or readability as measured by any of a number of different formulae. Response requirements can also vary. We can call these different responses by names like "understanding words in context," or "getting the main idea," or "understanding ideas expressed in more than one sentence." We can then investigate stimulus-response relations just like the animal psychologists do in their labs along the Charles River.

Comprehension tests of this type are quite limited in the extent to which they can tell a teacher what to do for individual students, even if they can provide information about relations between characteristics of text and types of response. Unfortunately, however, most tests of this type are not very useful. I would like to point out two reasons why they are not and offer some rather simple solutions.

One of the reasons is that these reading tests are constructed so that abilities which may have little to do with reading ability can have a great effect on test scores. The typical test construction procedure—the one suggested in all the textbooks—includes pretesting and item analysis. Perhaps half of the pretested items are discarded. The ones that are selected for the final test form are those which best predict total test scores of examinees. Yet total test score is influenced by inferential ability, general knowledge, test taking strategy and any number of human factors that one would be reluctant to call reading ability. The correct test construction procedure would require half of the pretest sample to attempt test questions after

reading the test passages; the other half of the pretest sample would attempt to answer the test questions after reading other, unrelated passages. The test assembler would then select items which discriminated between the two pretest groups.

I have actually reviewed professionally made tests of reading comprehension in which I, and others, could answer three quarters of the questions correctly without having looked at the test passages. These could hardly be called good reading tests. It is no wonder, therefore, that so-called reading tests of this sort yield scores which are little different from scores on other types of tests: listening tests or grammar tests for example. In fact inter-correlations among ESL tests are frequently so high that some authors have suggested doing away with everything but general proficiency tests.

This leads me to the other problem with contemporary reading tests. That is the <u>other</u> tests. I've shown that reading tests are not constructed so that they maximize abilities unique to reading; I'd like to suggest also that other tests do not <u>minimize</u> what is unique to reading. It is inevitable that there will be a positive relation between reading and other ESL test scores. But granted that a test situation is special and calls forth a regular form of behavior, regular stresses and strategies regardless of the kind of test; granted also that there are communalities across tests in knowledge of grammar and vocabulary and in the ability to utilize redundancy; granted also that some tests are hybrids—for example the listening test with written answer choices; granted also that in learning a second language most students will hear teachers and tapes and will have textbooks to look at, so that learning of different things tends to proceed apace—nonetheless differences which should exist do not.

Let me just illustrate.

The script of any listening test I've ever seen looks far different from a transcription of ordinary speech, especially when dialogues or informal extended discourse are used as stimulus material. You are surely familiar with the "cocktail party phenomenon," the ability of a listener to follow one person's speech when everyone in the room is talking. I've never seen that utilized in a listening test. In fact students and teachers complain violently if there is a little background noise on an examination tape. I haven't tried it yet, but I'm convinced that a "cocktail party listening test" would show a lower correlation with reading tests than do the listening tests we use today.

The modern reading comprehension test presents another problem, which is not, however, so easy to solve without going to different techniques. The format of power testing with multiple-choice questions is best handled by a student when he adopts a strategy designed to avoid false positives, that is to keep from being caught by an attractive distractor. This is accomplished when he sets a very high standard for information required before making a decision. This leads to slow reading, re-reading, and probably to translation; yet this is <u>not</u> the kind of skill we are primarily interested in. This certainly bears little relation to fluent reading.

Our contemporary reading tests can be improved but I think we can go further than that. I've long been convinced that the four skill model of language is a gross and not-very-useful oversimplification. But I am convinced also that just as people may be more or less effective speakers or readers in their own languages, so may they be more or less able readers in a second language.

It is assuming too much to say that a good reader in his own language will automatically be a good reader in a second language as soon as he learns a bit of grammar, a set of words and the alphabet. Some research I'm doing now tends to support this skepticism. I have learned that a group of ESL speakers who have learned the language in their work and use in in their work can deliver understandable spoken messages far better than a highly educated, intelligent group of university students with comparable measures of grammar, vocabulary, and pronunciation. It is reasonable to think that the ability to understand written messages is just as independent of grammar, vocabulary and letter recognition.

Now that we have established some sort of perspective for what is new, I have to report that my search for new reading tests hasn't turned up very much—not much, at least, that can be immediately incorporated into instructional and examination programs. Several things, however, must be mentioned.

John Bormuth's work On the Theory of Achievement Test Items (1970) has been getting a lot of attention lately. The theory he expounds is not yet worked out in sufficient detail to be applied and its purpose is directed towards measuring comprehension of a particular text, not towards measuring a student's ability to comprehend any text. Furthermore, what he offers is not so much a new testing method or new view of comprehension, but rather a way of improving the kinds of tests that are written today.

Ken Goodman's procedure for interpreting miscues during oral reading seems quite promising, (Goodman and Burke, 1969) but is only now being tried out with speakers of a second language. A miscue is any difference in what the reader says from what one would expect from the text. For example, the text reads, "He opened the volume to page fourteen," and the examinee says,

"He opened the book to page fourteen." In theory one should be able to attribute miscues and their spontaneous corrections to the student's grammatical knowledge, his semantic knowledge or his processing of visual information.

Oller's work in photography of eye movements of second language readers has important research interests but is not yet—as far as I know—a useable technique for large scale testing (Oller, 1972).

The cloze test—and one should include it's young cousin, the clozentropy test—is in the reading test tradition and has found new favor for testing second language abilities. Bosco's views on the importance of anticipation (Bosco, 1972) suggest a new version of the progressive cloze test, one in which students reading a text are periodically questioned about what themes will follow.

Those three techniques last mentioned—cloze, eye movement photograph and miscue analysis—have one important feature in common. They measure a smaller range of information processing tasks than do the contemporary reading comprehension tests. In reading one must first get the message, derive a meaning with the aid of a written text. After that one may evaluate the message for acceptability, he may incorporate it into his memory structure, he may use it to derive further cognitive structures on the basis of inference, and he may also retrieve the message from memory (Carroll, 1971). The current tests require an examinee to get the message, to incorporate it into memory and then to retrieve it in response to a test question. But comprehension is the process of getting the message. A test of reading comprehension should give information about that stage of the information processing sequence.

Reading comprehension depends upon the integration of information from

three different sources: visual information from print, permanent semantic information, and grammatical knowledge of the language (Smith, 1971). The reader with the greatest comprehension ability is, paradoxically, the one who uses the least amount of visual information.

Semantic information for speech or writing is probably the same. The "grammar" of written English—including the relative use of particular structures—clearly is not.

One might find out how well this integration has been accomplished by asking a student how well he understood a standard passage which he read in a given time. The reason we don't do this is probably that we won't trust our students. Surely lack of comprehension is a more prevalent problem than misunderstanding.

I haven't found any sound method for investigation the extent to which a reader relies upon each source of information. But let me suggest for your consideration a technique which might work. Of course the test contents must still be worked out.

Gradually increase the visibility of a random set of digits until an examinee can copy them. This establishes for him a measure of basic visual acuity. Then follow the same procedure with a standard test passage. The difference in required illumination is an index of reading comprehension ability. The greater the difference, the less visual information is being required and, therefore, the greater is reading comprehension ability. It might be possible to extend the technique so that examinees must respond to grammatical but anomalous displays, (for example: The book's orange dreams were watching my help.), and also that he respond to meaningful but ungrammatical sequences (rob thief yesterday mother jewelry). These measures

requiring more illumination than completely well-formed English text might index the relative contributions to comprehension of a students semantic and syntactic knowledge.

These last ideas are highly speculative, and we already know that close tests or passage-plus question tests can be easily improved and may be satisfactory for many purposes. But when we are concerned either with what is peculiar to reading or with the component processes of decoding written messages, much more complex and costly measures will be required. That is the price of asking more sophisticated questions. It is probably a price worth paying.

REFERENCES

Bormuth, John *On the Theory of Achievement Test Items.* Chicago: The University of Chicago Press, 1970.

Bosco, Frederick J. "The Reader's Expectancy of Old and New Information." Paper presented to the TESOL Convention, Washington, 1972.

Carroll, John B. *Learning from Verbal Discourse in Educational Media: a Review of the Literature.* Princeton: Educational Testing Service, 1971.

Goodman, Kenneth S. and Carolyn L. Burke *A Study of Oral Reading Miscues that Result in Grammatical Re-Transformations.* Final Report, Project No. 7-E-219, Washington: Office of Education, HEW, 1969.

Harris, David P. *Testing English as a Second Language.* New York: McGraw-Hill Book Co., 1969.

Lado, Robert Language Testing: *the Construction and Use of Foreign Language Tests.* London: Longmans, Green and Co., Ltd., 1961.

Oller, John "Assessing Competence in ESL: Reading" Paper presented to the TESOL Convention, Washington, 1972.

Smith, Frank *Understanding Reading* New York: Holt, Rinehart and Winston, Inc., 1971.

Valette, Rebecca M. *Modern Language Testing, a Handbook.* New York: Harcourt, Brace and World, Inc., 1967.